Data fusion of multi-view ultrasonic imaging for characterisation of large defects

Rhodri L.T. Bevan[1], Nicolas Budyn[1], Jie Zhang[1], Anthony J. Croxford[1], So Kitazawa[2] and Paul D. Wilcox[1]

***Abstract*—The multi-view total focusing method (TFM) enables a region of interest within a specimen to be imaged using different ray paths and wave mode combinations. For defects larger than the ultrasonic wavelength, different portions of the same defect may manifest in a number of views. For a crack, the tip diffraction response may be evident in certain views and the specular reflection in others. Accurate characterisation of large defects requires the information in multiple views to be combined. In this work, three data fusion methodologies are presented: a simple sum over all views, a sum weighted according to the inverse of the noise in each view and a matched filter approach. Four large defects are examined, one stress corrosion crack (SCC), two weld cracks and a pair of slagline defects in a weld. The matched filter (matched to a small circular void) provided significant improvement over the best individual view. The data fusion process incorporates artefact removal, where non-defect artefact signals within each image view are identified and masked, using a single defect-free dataset for training. The matched filter was able to accurately visualise the full 3D extent of the four defects, allowing characterisation via the decibel drop method. When compared to x-ray CT and micrograph data in the case of the SCC, the matched filter fusion provided excellent agreement. Its performance was also superior to any individual view while providing a single fused image that is easier for an operator to interpret than a set of multi-view images.**

***Index Terms*—Data fusion, matched filter, ultrasonic imaging, ultrasonic transducer arrays**

I. Introduction

The aim of any non-destructive evaluation (NDE) inspection is to reliably detect and accurately characterise defects within the inspected specimen. In an ultrasonic inspection, this can be achieved through use of a phased array system [1]. This is enabled in this work through the full matrix capture (FMC) approach [2], in which time-domain data from all possible combinations of transmitter and receiver in the phased array is stored. Post-processing of the time-domain data into a spatial ultrasonic image is then undertaken using synthetic focusing. A common algorithm to achieve this is the total focusing method (TFM) [3]–[5] with synthetic focusing on both transmission and reception using the linear delay-and-sum beamforming approach to achieve full focusing at each pixel within the image.

Use of mode conversions and ray paths involving reflections of specimen surfaces enables multi-view TFM images to be formed of the same physical region of interest (ROI) in a specimen from one FMC dataset [6]. Multi-mode reconstruction has also been incorporated in the plane-wave imaging algorithm [7], and extended to the Fourier domain in [8]. Modelling can be utilised to select the appropriate ray paths and modes to form the views [9]. Multi-view imaging provides more information, but human interpretation of these images is challenging. If a defect is visible in its entirety in a single view, the view in which this occurs depends on the location and type of defect. For large defects, different aspects of the defect response may manifest in different views: in some views, portions of specular reflection may be evident, while in other views only tip diffraction effects may be observed. In general, no one specific view can be expected to provide a complete picture of any defect. It is therefore desirable to apply some form of data fusion [10], [11] to combine information from multiple views into a single image. Data fusion enables the information that can be extracted to be maximised whilst minimising the data presented.

To avoid the data fusion procedure reducing resolution and detectability, it is essential to take into account multiple factors. It is necessary to recognise that the noise level is different between views. All the fusion methods considered here involve the combination of ultrasonic intensities from multiple views, and it is therefore also necessary to recognise that the expected signal amplitude from a defect is view-dependent. For this reason, information on both the expected noise and signal levels is needed. Micro-structural (grain) noise within a specimen is typically the limiting factor in detecting a defect hence significant effort has been placed into its characterisation. Early experimental work on the statistical properties of grain noise was undertaken by Wagner *et al.* [12]. Previous work [13] has provided an efficient and robust method of quantifying both random and coherent grain noise in multi-view TFM images. Expected signal intensities can be determined through the use of finite element modelling (FEM) or analytical models. Simulations must encompass defect responses at locations throughout the ROI as the location of real defects is not known *a priori*.

In this paper, the fusion of multi-view TFM images is examined with the specific aim of characterising large defects (defined as greater than 2λ, where λ denotes the longitudinal wavelength in the specimen at the nominal centre frequency of the ultrasonic inspection. The longitudinal wavelength was selected to ensure that the defect was large, even for the longest wavelength). Four large defects are examined within industrial steels exhibiting various levels of grain noise to provide a representative measure of the fusion methodology. The results

This work was supported by the RCNDE and EPSRC Grant No EP/N015924/1, together with industrial partners: EDF, Hitachi, BAE Systems and Wood plc.

[1]Department of Mechanical Engineering at the University of Bristol, Bristol, BS8 1TR, UK. Email: R.Bevan@Bristol.ac.uk

[2] Hitachi, Ltd. Research & Development Group 7-2-1 Omika-cho, Hitachi-shi, Ibaraki-ken 319-1221, Japan

obtained from one of these industrial samples are compared to both X-ray computed tomography and microphotographic images.

In the following section, the experimental framework and process for the creation of data for fusion is outlined. Section III details the procedure for data fusion. The remaining sections examine the accuracy of the approach for two industrial samples and explore how to apply it in practice to characterise the four defects therein.

II. Experimental framework and creation of data for fusion

A. Experimental configuration

In this paper, an immersed oblique inspection is considered as it is a common configuration for the inspection of welds in safety-critical components and it also enables multi-view approaches where more information can be extracted. This configuration is shown in figure 1, with the seven properties employed to quantitatively describe the experimental setup. The instrument delay, t_d, refers to the delay between the time origin of A-scans in raw FMC data and the time at which the centre of a transmitted pulse can be regarded as leaving a transmitting array element. It is assumed that t_d is uniform for all elements in the array. The couplant is water with ultrasonic velocity v_W.

Cartesian axes xyz are defined as shown in 1 with the z axis normal to the specimen surface. In this work, a 1D linear array is used and the x-axis direction is defined such that the array lies in the xz plane. Experimentally, it is necessary to ensure that the out-of-imaging-plane angle, i.e. the array angle relative to the y-axis, is zero. After this adjustment, the y-dimension of the array elements is parallel to the specimen surface. These constraints mean that the array orientation is completely defined by a single orientation angle, θ, relative to the x-axis as shown in the figure. The stand-off of the array is specified by Z_s, the distance of the centre of the closest element to the specimen surface. In this work, all specimens are assumed to be isotropic plates of thickness d, with longitudinal and shear ultrasonic velocities denoted by v_L and v_T respectively.

These properties were calculated after measuring the water temperature and material thickness. v_W was obtained from a published relationship between velocity and temperature [14], while v_L and v_T were calculated from the back surface signal response. For all experimental configurations, a 5 MHz, linear, 1D array (manufactured by Imasonic, Voray-sur-l'Ognon, France) with 128 elements and pitch of 0.3 mm was used. The element width in the xz plane was 0.2 mm and the element length in the y-direction was 15 mm. Data was acquired from the array using a array controller (MicroPulse 5 manufactured by Peak NDT, Derby, UK) at a sampling frequency of 25 MHz. The values of the seven parameters depend on the particular experiment and are presented in the relevant results in section IV.

To image in three dimensions with a 1D array, it is necessary to combine a sequence of 2D cross-sectional (xz plane) images. This y direction translation between cross-sectional locations was controlled via a motorised rig, capable of incremental steps of 0.0125 mm.

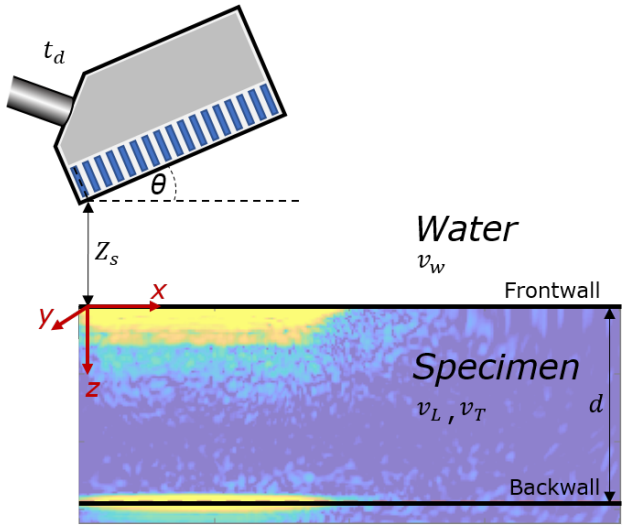

Fig. 1. Immersed oblique inspection setup

B. Creation of multi-view images

Prior to generating multi-view TFM images, the FMC data is filtered and Hilbert transformed in the frequency domain using a Gaussian window function (centred at the phased array centre frequency of 5MHz and -40dB half-bandwidth of 90% relative to the centre frequency). The TFM images are then generated using the summation of time-delayed data. The image for the i^{th} view is

$$I_i(\mathbf{r}) = \sum_{T=1}^{N}\sum_{R=1}^{N} a_{TR}^{(i)}(\mathbf{r})\tilde{f}_{TR}^{(i)}(\tau_{TR}(\mathbf{r})) \tag{1}$$

where $\mathbf{r}$ is the image position, $a_{TR}^{(i)}$ is an optional apodisation term and $\tilde{f}_{TR}$ is the post Hilbert-transformed FMC data. The subscript T denotes the transmitter element, R the receiver element and N is the number of elements in the array. The time delay $\tau_{TR}^{(i)}$ depends on the total travel time between transmitter T, $\mathbf{r}$ and the receiver R. The TFM algorithm, as given in Eq. 1, incorporates a weighting term, which can be employed to represent beam directivity. However, the standard TFM algorithm does not utilise this term, with each contribution treated equally, i.e. $a_{TR}^{(i)}(\mathbf{r}) = 1$. It is worth noting that with the immersed setup employed, the impact of beam directivity is minimal, with attenuation and beam spread dominating the spatial variation of response within the region of interest. This is accounted for using the spatial correction mentioned in Section II-C of the paper, with the detail provided in reference [13]. Lanczos interpolation (using the 3-lobed Lanczos-windowed sinc function) was used to interpolate the discretely-sampled FMC data [15] to the correct delay given by Eq. 1.

The multi-view TFM algorithm [6] utilises a ray-based model of wave propagation to generate different views of

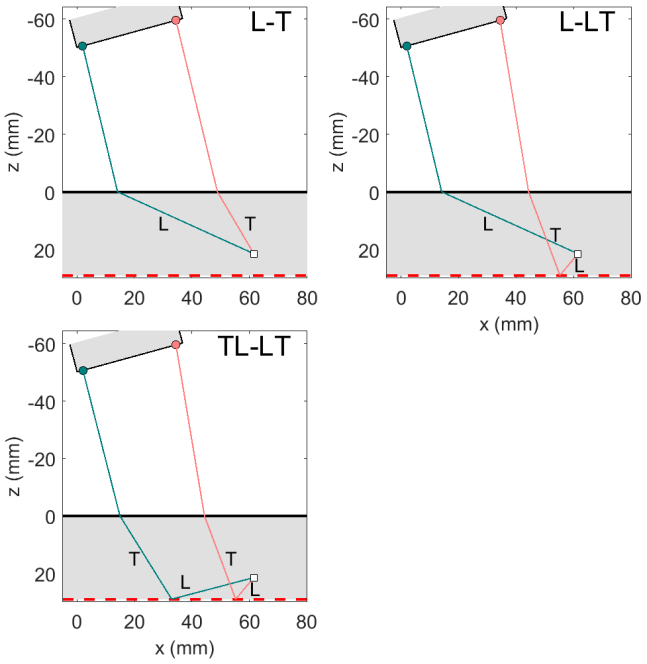

Fig. 2. Examples of direct (L-T), half-skip (L-LT) and full-skip (TL-LT) views. The view name denotes only the modes utilised within the specimen for each leg within the respective transmission or reception path.

the same physical region from a single FMC dataset. These views aim to utilise ray paths which maximise the viewing angle of any particular defect, making visible defect responses across these views more likely. By increasing the viewing angle range through considering multiple ray paths, this can avoid the need to physically relocate the array to achieve a similar viewing range. Not only does this increased viewing range make a small defect more likely to be seen, but it also enables a large defect to be examined from multiple angles, with features which may be invisible due to an insignificant reflection (or shadowing from a different part of the defect) in one view appearing in another. The terminology used to represent the different ray paths and wave modes (transverse, T, and longitudinal, L) is taken from [13], [16]. Views can be considered as direct, half-skip and full-skip, as shown in figure 2. A skip path denotes a single reflection off of the specimen backwall, either for the transmission ray path or reception ray path. A full-skip view is one whereby both the transmission and reception paths are via the backwall, whereas in half-skip views only 1 of the 2 paths have travelled via the backwall. For the dual medium setup considered, there are 6 potential outbound paths to arrive at a point within the specimen interior since a mode conversion (between T and L modes and vice versa) can occur at a solid/fluid boundary. Since the same holds true for the reception paths, this results in 36 combinations of transmission and reception, representing 36 views. However, due to reciprocity, only 21 are unique and only these 21 views are considered in this paper with the notation of [16]. A representative example of multi-view TFM images obtained from a single FMC dataset captured on a stainless steel sample (the full details of which are provided in section IV-A) is shown in figure 3. Two key points are evident from this example. Under the array (x <40 mm) many views are contaminated by artefacts. Artefacts are due to a signal response from another ray path appearing in the current view and reconstructing at the wrong point. They may occur throughout the image extent, but are most prevalent under the array in half-skip and full-skip views, where the direct path backwall reconstructs. Thus, any data fusion process using these views must consider artefacts as contamination to be avoided or eliminated otherwise the fusion algorithm will simply combine the noisy artefacts and pollute the information from cleaner views. The example presented in figure 3 contains a stress corrosion crack. This crack is V-shaped and located at approximately x =65 mm (see figure 4). The second point to observe from this example is that the defect response varies between views due to the ray paths for different modes being shadowed or visible. In the example views such as LT-T and TL-LT having a good response towards the leftmost fork whilst the direct views L-L and T-T have the strongest response to the rightmost fork. Indeed in T-T the response to the leftmost fork is barely detectable above the background noise. As such, no single view provides a complete picture of the defect, making the sizing of the defect from a single view problematic and highlighting the need for data fusion.

C. Creation of input data for fusion

Information-based data fusion requires knowledge of the signal-to-noise ratio (SNR) at every point in each multi-view image to merge the individual views in an advantageous manner. The noise level varies both between views and spatially within individual views, due to factors such as attenuation and beam spreading. An efficient procedure for the quantification of both random and grain noise was presented in [13] and is only briefly summarised here. Following the method outlined in that paper, random noise was found to be significantly lower in amplitude than coherent grain noise and is not discussed further. A single FMC dataset obtained from a pristine region of a component is utilised for grain noise quantification and artefact masking. A mask function for the i^{th} view, $m_i(\mathbf{r})$, is defined as 0 in artefact regions and 1 in artefact-free regions. In a pristine specimen, the image in the $m_i(\mathbf{r})$ =1 region should contain only grain noise, and this is the hypothesis behind the automated process to identify artefacts and characterise grain noise. The basic principle (applied separately to each view) is the iterative growth of the $m_i(\mathbf{r})$ =0 region until the statistical properties of the image in the $m_i(\mathbf{r})$ =1 region are those expected from speckle noise. The spatial variation in mean speckle amplitude within each view, which is predominantly caused by attenuation and beam spreading, is approximated by a log-linear function of position. At each iteration, the image intensities in the $m_i(\mathbf{r})$ =1 region are fitted to such a function and then multiplied by its reciprocal, $c_i(\mathbf{r})$. If no artefacts are present, the result of this multiplication should be speckle with a Rayleigh distribution of intensity that is spatially uniform with Rayleigh parameter σ_i; if this is not the case, the $m_i(\mathbf{r})$ =0 region is expanded to include the highest normalised intensities and the process repeated. Once $c_i(\mathbf{r})$ and $m_i(\mathbf{r})$ have been determined from the pristine dataset, the

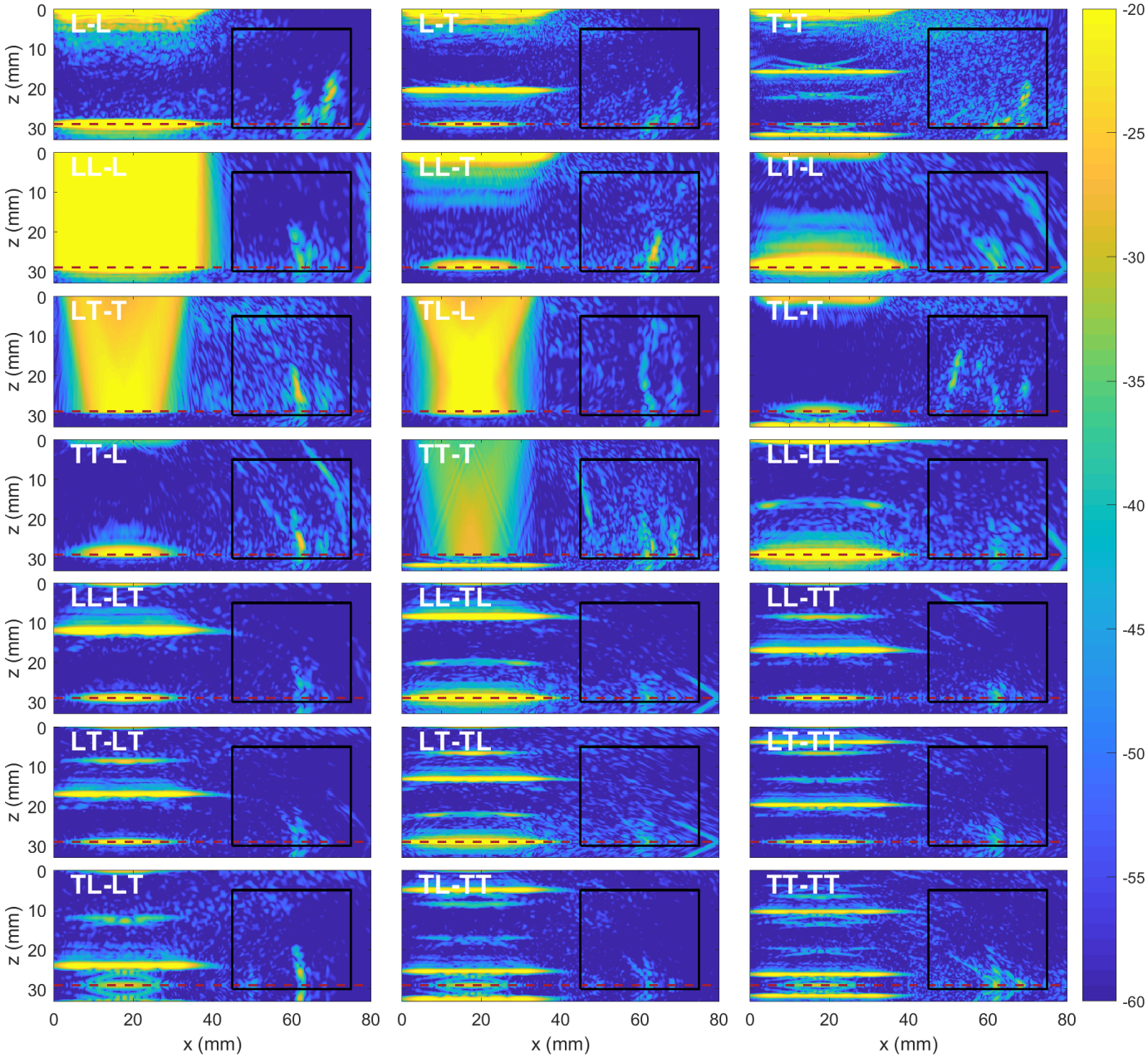

Fig. 3. Experimental example of the 21 unique TFM views (in dB) for a stainless steel sample containing a stress corrosion crack, considering only direct, half-skip and full-skip ray paths. The backwall is denoted by the red dashed line in each view. The defect is located at approximately $x = 65$mm and the black box outlines the region of interest for this sample.

masked and corrected images for subsequent measurements are calculated as $x_i(\mathbf{r}) = c_i(\mathbf{r})m_i(\mathbf{r})\,|I_i(\mathbf{r})|$. The number of views contributing data to each point is also recorded as $n(\mathbf{r}) = \sum_i m_i(\mathbf{r})$.

In addition to the noise level, the expected signal level in each view is also required. Thus it is necessary to model a representative defect to achieve a suitable expected signal intensity. Since TFM image reconstruction is undertaken in 2D, only 2D defects (infinitely long in the out-of-plane direction) are considered. This defect response must be modelled at a number of positions within the imaging domain so that a predictable response intensity at all points in the image can be achieved. Whilst signal prediction can be targeted towards a specific defect type, this presents two difficulties. It can significantly increase the complexity of the problem, for instance a crack-like defect, having a particular length and orientation, would add two additional parameters to be investigated, along with its location. Secondly, if a very specific defect is considered then the sensitivity to alternative types of defects will be potentially reduced, especially when defects are large.

Defect modelling can be achieved using analytical models such as the commercial software CIVA [17]. Here, a single-frequency ray-based model previously developed by the current authors is used [16]. This 2D based model incorporates directivity, transmission & reflection coefficients and beam-

spread along with attenuation. The attenuation coefficients are determined from the experimental data for each sample. The model is used to produce sensitivity maps, $E_i(\mathbf{r})$, for each view, i, that show the expected amplitude response to a specified type of defect. Note that the sensitivity output by the model is scaled within each view by $c_i(\mathbf{r})$ so that it is consistent with the experimental data, $x_i(\mathbf{r})$. It was shown that this approximate model predicted the peak amplitude of the response of small defects to within ± 3 dB of those measured experimentally or predicted by a multi-frequency model (requiring 3 orders of magnitude greater computational load). The defect modelled for the sensitivity map used here is a very small side-drilled hole (SDH) with a radius of 0.001 mm. The implicit assumption behind this choice is that larger defects can be considered as the superposition of the responses from many such holes. The analytical model utilised to construct the scattering matrices for a SDH is outlined in [18]. The use of a scattering matrix corresponding to a small physical scatterer ensures appropriate scaling between the four different inbound-outbound modalities (i.e. LL, LT, TL, TT).

The inputs to the data fusion process are $E_i(\mathbf{r})$, $n(\mathbf{r})$, σ_i and either $I_i(\mathbf{r})$ or $x_i(\mathbf{r})$ depending on the method.

III. Data Fusion

With an understanding of the predicted signal intensities, removal of artefacts and grain noise quantified, different fusion approaches may be investigated. From a single FMC dataset, 21 multi-view images are obtained. To examine the effect of the fusion algorithm on combining these 21 views into a single fused image, three approaches are considered. The first and most basic fusion approach is to simply take the average of the uncorrected image intensity values using

$$I_B(\mathbf{r}) = \frac{1}{N_v} \sum_i |I_i(\mathbf{r})| \,. \tag{2}$$

where N_v =21 is the total number of views. Two informed fusion methods are also considered, firstly fusion based on equalisation of the noise levels between images and the second is a matched filter tuned to the representative defect response.

The noise-weighted fusion approach is defined as

$$I_W(\mathbf{r}) = \frac{1}{n(\mathbf{r})} \sum_i \frac{x_i(\mathbf{r})}{\sigma_i}. \tag{3}$$

where the division by $n(\mathbf{r})$ ensures a consistent amplitude because the number of views with unmasked regions that contribute to the summation varies by pixel.

The matched filter approach correlates the representative defect response, $E_i(\mathbf{r})$, to the measurement, $x_i(\mathbf{r})$, in order to detect the presence of the former within the latter. This approach is the optimal linear filter for maximising the SNR. The matched filter fusion approach is given by

$$I_M(\mathbf{r}) = \frac{1}{n(\mathbf{r})} \sum_i \frac{x_i(\mathbf{r}) E_i(\mathbf{r})}{\sigma_i^2}. \tag{4}$$

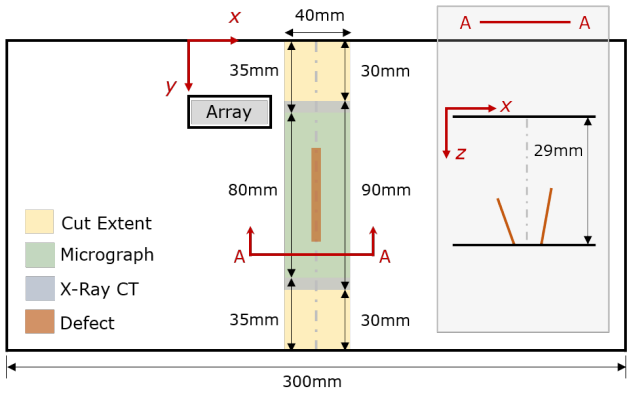

Fig. 4. Sample A schematic showing the corroborating data collection zones surrounding the V-shaped crack. The centreline is located at $x = 65$mm relative to the reference element.

TABLE I
SAMPLE A: EXPERIMENTAL SETUP PROPERTIES.

Probe vertical standoff, Z_s	50.2 mm
Probe inclination angle, θ	15.3°
Water velocity, v_W	1485 m/s
Instrument delay, t_d	580 ns
Specimen thickness, d	29.0 mm
Specimen longitudinal velocity, v_L	5833 m/s
Specimen shear velocity, v_T	3181 m/s
Specimen longitudinal attenuation, α_L	2 Np/m
Specimen shear attenuation, α_L	8 Np/m

IV. Results

In this section, the three fusion approaches described are applied to two industrial samples containing large (greater than 2λ) defects. Sample A is a stainless steel plate containing a stress corrosion crack which has also been investigated using X-ray computed tomography (CT) and microphotographic observation. Sample B is a ferritic steel plate with a weld containing three manufactured defects of specified dimensions (two cracks and a pair of slag lines).

A. Sample A: Stress Corrosion Crack

The stress corrosion crack (SCC) was grown in a JIS SUS304 austenitic 150×300×29 mm thick stainless steel plate. A schematic is provided in figure 4, showing the defect location within the sample. For comparison analysis, the sample was reduced in size as shown in the figure (by the grey cut extent) before inspection with an X-ray CT machine which obtained sections along a 90mm region. Finally, destructive, micrographs were obtained along an 80 mm region, with 21 cross-sections at an interval spacing of 4 mm. During ultrasonic data collection, the phased array was located to the left-hand side of the sample centreline (illustrated in figure 4 as Array), with the centreline occurring at $x = 65$mm relative to the reference element. The experimental setup parameters are provided in table I. 84 FMC datasets were collected along a 126 mm traverse in the y-direction in 1.5 mm increments. This data collection took 90 minutes.

A representative set of multi-view TFM images for this sample was shown previously in figure 3. In these images it is

TABLE II
GRAIN NOISE LEVELS PER VIEW (IN DB), NORMALISED BY THE MAXIMUM LEVEL WITHIN EACH RESPECTIVE SAMPLE.

View	Sample A	Sample B
L-L	-9.5	-6.2
L-T	-8.0	-9.8
T-T	0.0	-3.7
LL-L	-8.6	-5.6
LL-T	-4.0	-4.1
LT-L	-6.6	-4.7
LT-T	-2.4	0.0
TL-L	-8.8	-6.2
TL-T	-7.4	-5.3
TT-L	-6.7	-7.4
TT-T	-3.2	-5.1
LL-LL	-5.8	-5.8
LL-LT	-12.2	-11.6
LL-TL	-9.3	-9.2
LL-TT	-11.3	-12.0
LT-LT	-11.8	-10.2
LT-TL	-6.2	-6.0
LT-TT	-9.1	-10.1
TL-LT	-10.0	-8.3
TL-TT	-11.9	-10.9
TT-TT	-7.3	-8.5

clear that getting an unambiguous understanding of the defect extent is challenging. For the results presented below, the ROI in the xz-plane extends from x =[45,75] mm (or [-20, 10] mm with respect to the sample centreline) and z =[5,30] mm with a pixel size of 0.05 mm.

In order to determine the grain noise parameters, a single FMC dataset from a pristine defect-free region was utilised to determine the spatial variation correction and the Rayleigh parameter per view. This parameter is used to assess the relative grain noise between each view, and is presented (in dB) in table II for both Sample A and B.

Using a density of 7850 kg/m^3, an equivalent sensitivity map is constructed for the 0.001 mm radius SDH defect within the region of interest and is shown in figure 5. In the figure, it is evident that the T-T view has the strongest expected signal intensities, although from table II, this is also the view with the highest noise level.

With the noise and expected signal sensitivity determined, the three data fusion approaches are now examined. The fusion of the 21 multi-view TFM images are shown in figure 6. From the figure, it can be seen that due to lack of significant artefacts in the ROI, the basic fusion method produces an acceptable result. If the ROI was extended this would not necessarily be true. In the basic data fusion method, the right-hand fork of the V-shaped SCC is less well defined than the left-hand one. The same is true in the case of the noise-weighted fusion algorithm. However, the matched filter fusion algorithm demonstrates a clear improvement, with both forks displaying similar intensities. The SNR of the matched filter is also better than the other two fusion approaches as shown in figure 6e. The L-L view is the best individual view for this particular defect. It has a better SNR than any of the fused images but does not provide good definition of the left fork of the SCC, highlighting that a single view is poor for characterising large defect extent. Also, the selection of the L-L view requires it to be manually identified as the best view. All three of the fusion algorithms can be adversely impacted by defect-based artefacts, as shown in figure 6e, which has resulted in a 3rd peak at $x = 52$mm. This defect-based artefact is most evident in the corresponding region of the basic fusion method (figure 6a). The right-fork artefact originates from the TL-T view, with the weaker left-fork artefact originating from the TT-T view. The matched filter has performed better by suppressing the left-fork artefact of the V-shaped SCC. However, the intensity of the bottom right fork in the matched filter case compared to the other fusion methods is lower. This is due to some relevant defect information being masked. Masking is also evident in the top left corner of figure 6d. Thus, where the artefact suppression corresponds to a defect location, the fusion process relies on other views providing sufficient information. Also, if the number of views is limited in part of the image, the masking transition may be evident. As the results in figure 6 are representative for the three fusion algorithms, only the matched filter is examined hereafter.

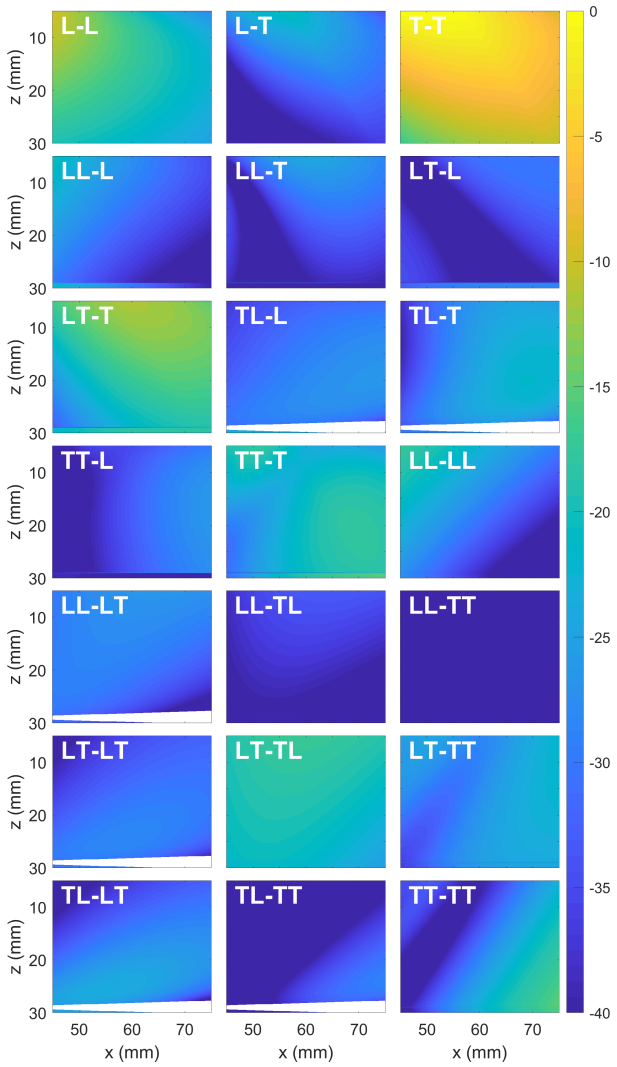

Fig. 5. Sample A sensitivity map (in dB) for a 0.001 mm radius SDH. These expected value maps have been scaled by $c_i(\mathbf{r})$. Values are normalised to the overall maximum scaled sensitivity.

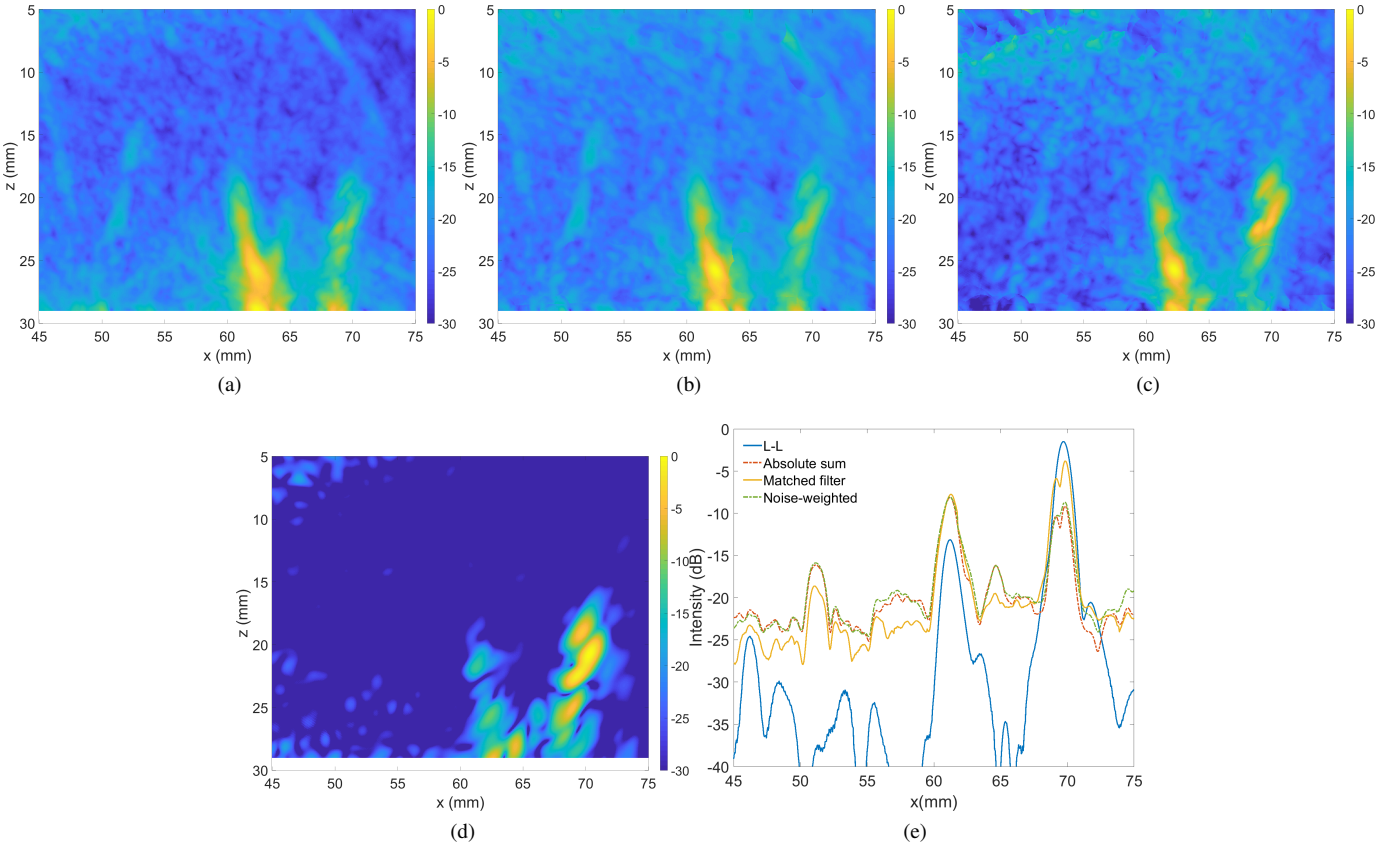

Fig. 6. Data fusion of the multi-view TFMs given in figure 3. (a) absolute sum (b) absolute noise weighted sum (c) matched filter (d) L-L view (e) Intensity profiles along z =22 mm. For (a)-(d) Each image is normalised by the maximum intensity occurring in that image.

A representative subset of cross-sectional micrograph observations of the V-shaped crack are presented in figure 7, together with the corresponding ultrasonic matched filter fusion and L-L view images. The ultrasound data has been normalised based on the maximum intensity present in the respective 2D cross-section. The X-ray data for the same locations are also included. For y locations where the cross-sectional micrograph lies between ultrasonic data points, Lanczos interpolation is again utilised to interpolate the ultrasonic data. The datasets from the three modalities were approximately co-registered by manually translating each dataset as a whole until a reasonable match was obtained. From the figure, it is evident that both the X-ray CT and ultrasound are managing to qualitatively image the full crack extent, at least for the matched filter fused images. The leftmost fork of the V-shaped crack is unreliably detected with the L-L view. The X-ray CT data has been processed to improve defect visibility. A moving average (in the y direction) was utilised to reduce noise, with a window size of ± 1 mm. Finally, the pixel intensities were normalised by their average value when no defect was present in the image, to remove spatial variation within the image. This average value was determined by using the initial and final 5mm of data, where no defect was visible. Qualitatively, the X-ray CT data is still extremely noisy, making determination of the crack extent challenging at best. From the figure, the L-L view is clearly outperformed by the matched filter and it is therefore no longer considered suitable for quantitative analysis of this sample. This clearly demonstrates the benefit of fusing all views and not relying on a single best view. The crack sizes, measured ultrasonically using the dB drop methodology [19] for selected dB thresholds is presented for the matched filter in table III. To determine the crack extent for each threshold, the left and right forks are analysed independently. To separate the two forks, the data is split at the midpoint, approximately $x = 65$ mm. Prior to splitting, the intensity values are normalised per image, thus treating each 2D slice as an independent snapshot of the complete crack. For each fork, the smallest possible rectangular box in a certain orientation is then fitted that encloses values exceeding the threshold (relative to the overall image peak) in the fork. This is repeated as the box orientation of the box is changed. The length of the longest box across all angles is used as the measure of fork length. From the table, it is apparent that both the -12 dB and -18 dB thresholds can provide good agreement. The root-mean-square (RMS) error, compared to the micrograph data, is 6.5, 3.3 and 3.4 mm for the -6, -12 and -18 dB thresholds respectively. However, whilst this absolute error metric is similar for the -12 and -18 dB thresholds, it is necessary to consider their respective bias and scatter. The standard deviation error (relative to the micrograph data) is 3.4 mm for the -12 dB threshold and 1.7 mm for the -18 dB threshold. There is no significant bias for the -12 dB threshold, with a mean error of 0.2 mm. The -18 dB threshold has a mean of 3.0 mm, displaying an

TABLE III
SCC DEFECT CHARACTERISATION, USING THE MICROGRAPH (MG) FOR COMPARISON BETWEEN THE RESULTS OBTAINED FROM THE MATCHED FILTER.

	Left fork (mm)				Right fork (mm)			
Slice	MG	-6 dB	-12 dB	-18 dB	MG	-6 dB	-12 dB	-18 dB
B	5.9	0.5	5.7	8.8	6.8	9.1	10.7	11.9
E	11.2	7.4	12.4	13.5	12.3	11.6	13.3	15.4
G	12.2	11.7	13.7	17.3	13.5	13.2	16.1	17.3
I	10.9	6.4	10.1	14.2	12.9	14.4	16.3	17.0
M	10.8	1.2	7.5	10.9	15.2	0.0	15.2	16.1
O	11.7	5.5	11.9	13.1	14.1	9.8	16.0	17.1
Q	10.5	1.1	5.1	11.5	11.3	0.0	3.0	13.0
T	5.5	5.3	6.7	9.1	4.7	0.5	8.9	11.0

oversizing bias. However, the micrograph data itself cannot be seen as a ground truth, as it likely underestimates the crack extent. Therefore, since some apparent conservativism would be preferential from a structural integrity perspective to an increased scatter and undersizing (e.g. slice Q), the -18 dB threshold provides the best compromise. Decreased characterisation accuracy is observed at the crack extremities (along y axis). This is believed to be due to the 15 mm length of the array elements in the y-direction; the size of the crack in the xz plane varies rapidly in the y-direction near its extremities resulting in blurring of the response in the ultrasonic imaging plane and inaccurate sizing. Still generally excellent performance is seen from the fused ultrasound data.

B. Sample B

The second specimen, Sample B, is an approximately 40 mm thick ferritic steel plate, containing a double V weld with three embedded defects. A drawing of the defect locations is shown in figure 8. Defects a and c are cracks in the weld, whilst defect b is a pair of slaglines, one above the other.

The surface of the double V weld has been dressed to level it, relative to the overall plate surface, in an attempt to produce a smooth continuous flat frontwall and backwall. However, ultrasonically significant surface variation is still present, with local deviation present in the block close to the weld. In practice, real components may not have perfectly flat, planar surfaces and it may not be permissible to machine flat surfaces to aid inspection. Sample B is inspected in its as-supplied state and is used to demonstrate that the procedure can adapted to accommodate non-planar surfaces. As neither surface can be adequately modelled as a flat plane, it is necessary to determine the actual surface profiles prior to their use in multi-view imaging. As for Sample A, the specimen properties are determined in a region of the plate where the front and back surfaces are parallel with a known (measured) thickness between surfaces. These experimental setup properties are provided in table IV.

The procedure for curved surface determination is as follows. The front surface is extracted from a single-medium TFM image generated based on the water velocity v_W, with the image covering the region immediately below the array. At each x-position within the TFM image, the maximum amplitude response from a 1D convolution kernel is calculated and

TABLE IV
SAMPLE B: EXPERIMENTAL SETUP PROPERTIES

Probe vertical standoff, Z_s	50.1 mm
Probe inclination angle, θ	12.8°
Water velocity, v_W	1473 m/s
Instrument delay, t_d	425 ns
Specimen thickness, d	40 mm
Specimen longitudinal velocity, v_L	5877 m/s
Specimen shear velocity, v_T	3227 m/s
Specimen longitudinal attenuation, α_L	3 Np/m
Specimen shear attenuation, α_T	8 Np/m

denotes a potential frontwall location at that x-position. The triangular kernel is utilised as a smoothing operator, producing an average weighted response along the z-axis. Using this kernel approach provides control over the level of smoothing and locality, through use of a single parameter, namely the length of the filter. For a z location z_0, with a support radius of 0.1mm, the kernel is defined by $K(\mathbf{u}) = (1 - \|\mathbf{u}\|)$, with bounded support for $\|\mathbf{u}\| < 1$ and $\mathbf{u} = z(\mathbf{r}) - z_0)$. Spurious surface points are removed by only considering those with an amplitude within 12dB of the image peak intensity. This is intended to remove points which have a limited signal response due to either the surface geometry or element directivity. In the single-medium images of the front surface, the TFM resolution in the x and z-directions were 0.05 mm and 0.025 mm respectively, with the increased z resolution to provide improved accuracy whilst minimising computational cost. A traverse of the sample in the x direction provides a series of (possibly overlapping) potential front surface points. These surface points are combined with a smoothing function based on robust local regression (RLOESS) with a 10% span. To ensure good coverage, the overlap between consecutive imaging positions was no less than 50%. A similar method has been presented in [20]. With the front surface defined, the back surface is now characterised in a similar manner, with the single-medium TFM replaced by the longitudinal direct mode TFM image (L-L) calculated using the measured front surface position. As before, only the region under the array is imaged at each location. An example of this procedure is illustrated in figure 9 for the front surface. For 3D imaging, data was collected at an interval of y =1.5 mm along a 270 mm scan traverse. The 3D surfaces can be seen in figure 10. From the figure, although each y position is extracted independently, agreement is evident between consecutive positions for both surfaces, suggesting the process is working effectively. The ultrasonically measured thickness of Sample B was found to deviate from -1 mm and +0.2 mm to the nominal thickness, d, within the region considered.

Due to the high-resolution in y and the secondary traverse in x for surface extraction, the total number of FMC datasets captured for Sample B exceeded 3,500. At each position, two FMC datasets were obtained, using adaptive gain to avoid over-saturation of the front surface and round-off error for the back surface. In total, data collection required 34 hours, therefore an additional FMC dataset was periodically acquired every 15 mm along y traverse with a known Δz =10 mm to monitor the speed of sound changes in water due to

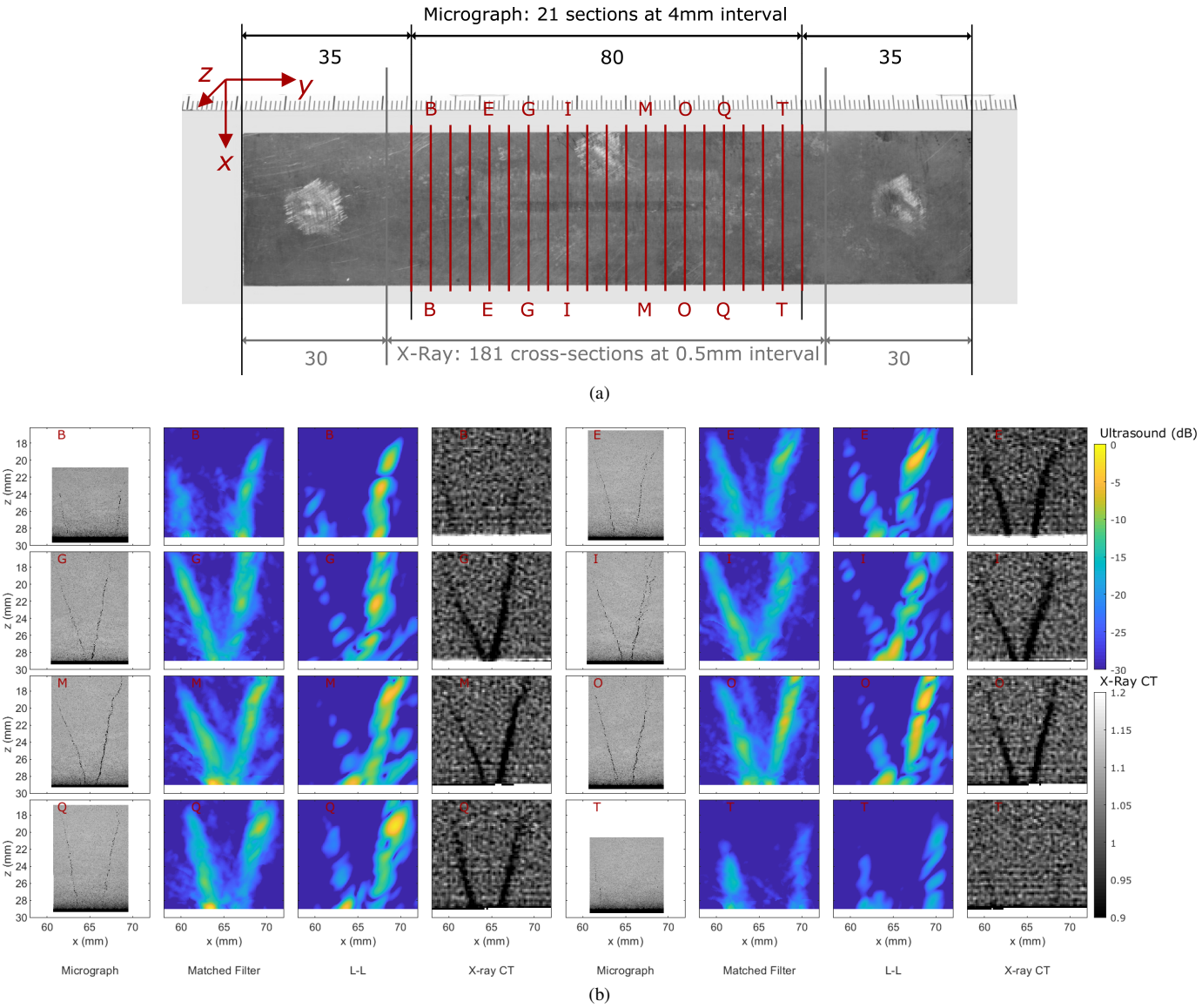

Fig. 7. Qualitative comparison between the matched filter fusion algorithm, the single L-L view and the X-ray CT and micrograph data at selected cross-sectional slices within Sample A (a) Photo of the reduced Sample A, with relevant coverage of collaborating data overlaid. (b) Selected cross-section slices, with their locations indicated in (a).

temperature. This velocity varied between 1467-1473 m/s over the collection period, with linear interpolation between monitored points used in the 3D reconstruction.

The relative grain noise parameters obtained from a single FMC dataset on a pristine, defect-free region are presented in table II for a ROI of x =[80,120] mm and z =[2, 37.4] mm. A representative set of multi-view TFM images from one FMC dataset is presented in figure 11 for y =57 mm with defect crack a present. As for Sample A, the L-L view appears to provide a reasonable image of the full extent of the crack, although indications from the crack are present in the majority of views. Again, understanding the extent of the defect by manually interpreting information from multiple views is challenging.

The interior of Sample B is shown in figure 12. In the figure two images are presented. The first is the fused view obtained from the matched filter algorithm, and the second is the L-L view. Both images are normalised to their respective peak defect intensity within the 3D ROI, and three isosurfaces at 6dB intervals are utilised to visualise the interior (i.e. -6,-12 and -18 dB). From the figure, it is evident that the three defects would have all been detected using the matched filter fusion algorithm, while in the L-L view the lower slagline is not detected. For defect a, the matched filter has excellent visualisation of the crack whilst the L-L view has a less obvious response which could be misinterpreted as a smaller void like feature rather than a crack. For defect c, the lower tip diffraction response dominates, with the expected specular response not evident in either the matched filter result or the L-L view.

By examining the three defects in isolation, and normalising based on each local defect amplitude, as shown in figure 13 for a representative vertical cross-section of each defect, it is possible to clearly see the improvement that the matched

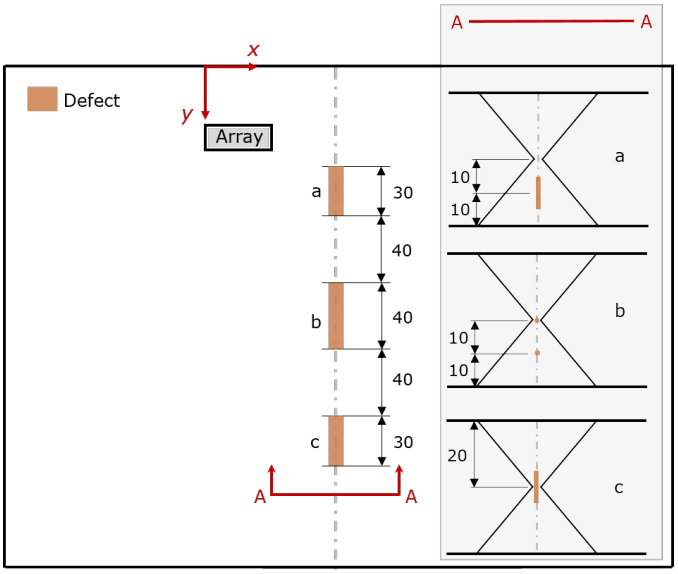

Fig. 8. Schematic of Sample B, a ferritic block containing three defects. Defects a and c are vertical cracks 8-10mm in length and defect b is twin slaglines

TABLE V
SAMPLE B DEFECT CHARACTERISATION: SIZE (MM) IN xz PLANE. FOR DEFECT b, THE SIZE IS THE CENTROID TO CENTROID SPACING BETWEEN THE TWIN SLAGLINES IN THE xz PLANE. EXPECTED (E) VALUES ARE TAKEN FROM FIGURE 8.

		L-L			Matched Filter		
Defect	E	-6 dB	-12 dB	-18 dB	-6 dB	-12 dB	-18 dB
a	10	2.8	3.7	4.9	2.9	3.7	10.4
b	10	-	-	12.0	10.5	10.9	11.1
c	10	2.0	2.9	4.4	1.8	2.7	4.5

TABLE VI
SAMPLE B DEFECT CHARACTERISATION: DEPTH OF DEFECT CENTROID (MM). FOR DEFECT b DEPTHS ARE REPORTED FOR THE TWO SLAGLINES (UPPER, b_U, AND LOWER, b_L). EXPECTED (E) VALUES ARE TAKEN FROM FIGURE 8.

		L-L			Matched Filter		
Defect	E	-6 dB	-12 dB	-18 dB	-6 dB	-12 dB	-18 dB
a	25.5	25.5	25.5	25.7	25.6	25.6	28.8
b_U	20.0	20.9	20.9	20.8	20.8	20.6	20.6
b_L	30.0	-	-	32.8	31.3	31.5	31.6
c	20.0	20.9	20.9	21.4	20.6	20.7	21.4

filter provides. In two of the three defects, the matched filter outperforms the L-L view by improving the signal response along the full defect profile. From the figure, the profile for defect a has typically 6-10 dB improvement whilst for defect b the lower slagline peak intensity has been improved by 15 dB. The performance improvement of the matched filter fusion for defect c is negligible, but equally importantly it is not degraded. This is due to other views making a limited contributions to the final image as a result of their poor SNR. These other views however do have some backwall artefacts which are not masked as the back surface has differed sufficiently from the training dataset to render the masking less effective. This is responsible for the minor increase in noise present in the matched filter beyond z =25 mm.

Characterisation of the crack length in the xz plane, based on the dB drop method, is impacted by the strong diffraction response. In the case of defect a elevated signal responses occur along this crack length. For defect c, the signal response is more dependent upon the chosen y location, as evidenced in figure 12. The true nature of these manufactured defects, especially the two cracks, is unknown. However, because their tip diffraction responses in direct views are strong compared to their specular responses in half skip views, it is believed that they are rough. This again highlights the additional insights that the data fusion can provide on the nature of large defects.

Quantitative characterisation of the three defects is presented in tables V and VI for the size and centroid location respectively when compared to the specified manufactured values as given in figure 8. These properties are determined in the same manner as for Sample A, with the rectangular box fitted around the values exceeding the specified threshold. The centroid of this box is used for the calculated defect centroid. From table VI, there is very good agreement between the measured defect centroid location and its expected position. The twin slagline locations have been examined by their constituent parts. From table V the 18 dB drop best corresponds to the defect size for defect a for the matched filter. As noted previously, the L-L view does poorly in sizing and detecting defects a and b. Rather than sizing the slaglines, which are of unknown size, the centre-to-centre spacing is reported. The matched filter shows excellent agreement with the expected value of 10mm for all three dB values, although accuracy decreases marginally as the dB value decreases as the extent of the slagline is no longer captured as cleanly.

V. DISCUSSION

Two industrial samples, with four defects, have been studied to examine the advantages offered by data fusion in characterising large defects. The matched filter fusion method, with its generic targetting of defects through use of a very small SDH defect for the expected signal, has shown that the combining of multi-view TFM images can significantly improve the quality and ease of visualisation of the whole defect profile. Obviously, as this is aiming generically, there is further potential for improving the matched filter response by tailoring the fusion process to the expected defect. However, an additional complication would then be introduced. Consider a non-circular defect (say a generic short crack segment), for each orientation angle used an expected defect, would result in a different fused image. This would then require a strategy to fuse the fused images for each orientation to obtain a single image.

The time required to undertake the fusion component in this current generic case is trivial (less than 1%) compared to the current rate-determining step which is the time taken to produce 21 TFM images from one frame of FMC data. This assumes that the sensitivity maps can be produced *a priori*. Otherwise, the sensitivity images require the same computation as TFM images, introducing an additional burden. However, since an inspection is typically specified in advance, this would not normally be an issue.

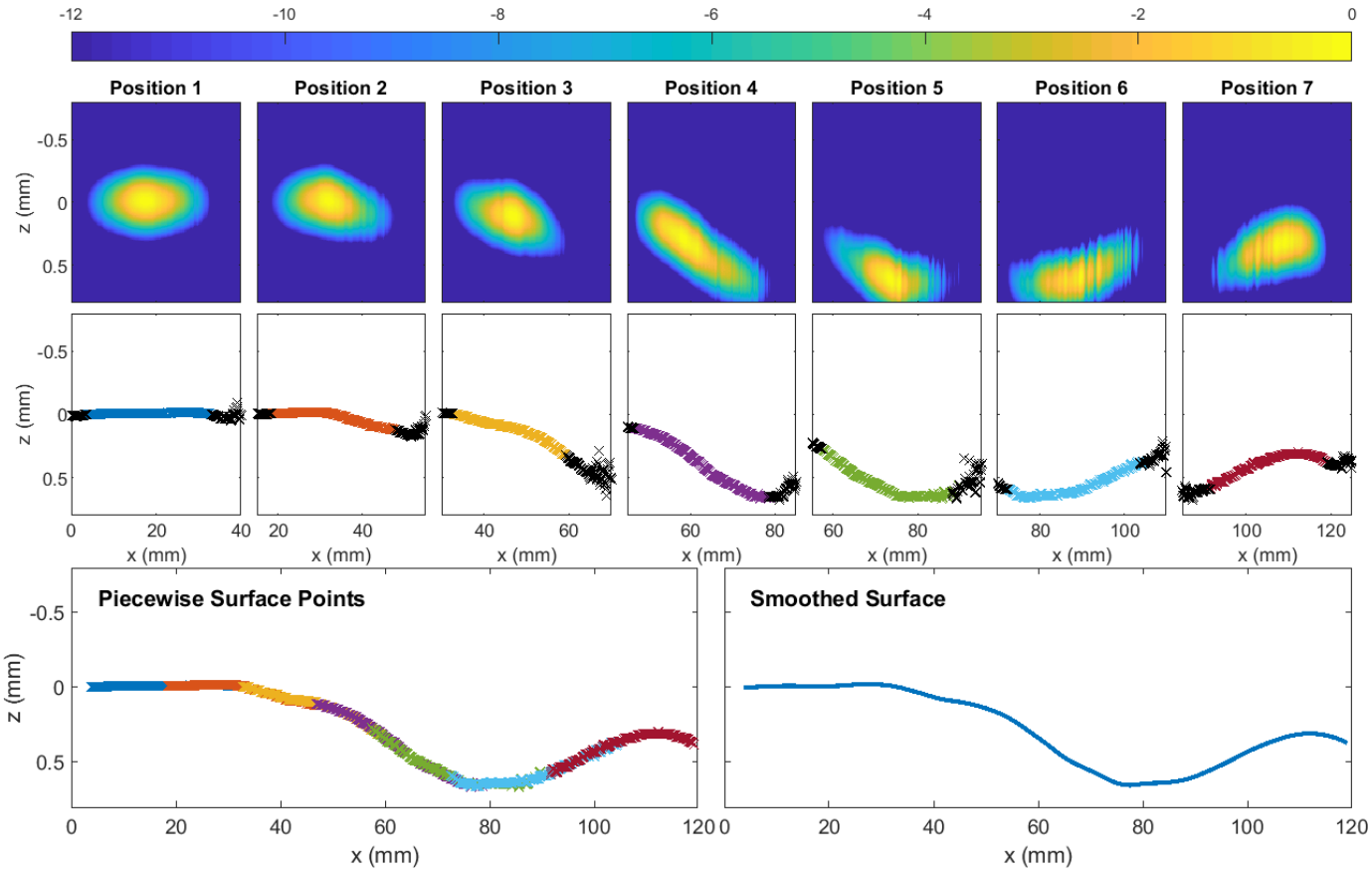

Fig. 9. Front surface extraction procedure example for 7 locations in an x axis traverse. Array offset for these 7 locations was Δx = [0, 15, 30, 45, 55, 70, 85] mm. The first row shows the single medium TFM image for each array position and the second row shows the extracted front surface coordinate at each x position. The black crosses denote points that are discarded because their amplitude is <-12 dB. The third row shows the combined surface.

The noise quantification procedure described in section II-C assumes that the material microstructure is spatially uniform. In the case of Sample B, the ROI contains both parent and weld material, which have different microstructures leading to different image noise levels. This effect is visible in the T-T view in figure 11, where the weld region has lower noise than the parent material. If necessary, a more sophisticated noise estimate can be made that takes into account known material heterogeneity. In this particular example where all defects are in the lower noise region, the measurable effect of doing this is negligible, but it would be potentially significant if defects in both parent and weld material were present in the sample.

Noise quantification also relies on the use of a representative defect free dataset(s). Only a single dataset has been employed in this work, although Sample B with its varying specimen surfaces may have benefited from updating of the mask. This surface variation also impacts on the sensitivity map. For the results presented, the sensitivity map was constructed once, at the same location as the noise calibration sample. Sensitivity maps can be readily constructed for each local 2D surface profile. However, since these sensitivity functions are smooth and slowly varying, a slight change in geometry is not going to make much difference. Whilst this was investigated, the 3% thickness variation in Sample B had a negligible impact on the results and is therefore not presented. With the mask, the slight geometry changes result in expected artefact locations varying slightly. A practical solution to this issue is to measure/estimate the expected deviations in specimen geometry and then dilate the mask to accommodate the resulting expected variation in artefact positions.

Image registration plays a key role in the abilities of a data fusion algorithm to combine relevant data in a manner which improves detection capabilities. Whilst this work has assumed that the image registration has been achieved to an acceptable degree through ultrasonically measuring variables such as the specimen velocities, it is possible to account for image registration issues through combining nearby pixel information into resolution elements (resels). This would degrade the apparent fused image resolution, although that resolution is illusory in situations of poor registration. Combining pixels into resels should only enhance detection although it may degrade overall SNR [11], leading to decreased overall detection performance (e.g. worse probability of false alarm for given probability of detection or vice versa). This resel combining could be an additional step when image registration is less than ideal.

Since characterisation of the defects is the focus in this work, the data fusion approach here has focused on image-based filtering. An alternative probabilistic approach is one whereby the TFM images are converted to probability values (p-values) prior to fusing. Using the statistical noise distribution (section II-C), signal-like intensities are then examined

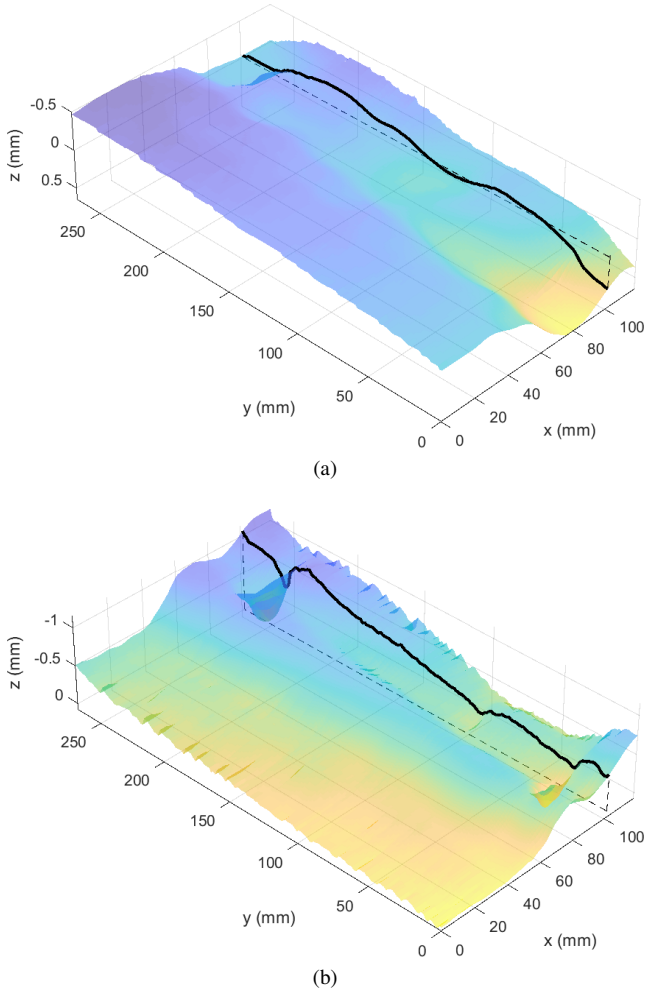

Fig. 10. Extracted 3D surfaces along $y = 270$mm traverse: (a) front wall; (b) back wall. The black line denotes the centreline of the double V weld along which the defects are located relative to $z = 0$ mm in (a) and $z = d$ in (b) using the black dashed line.

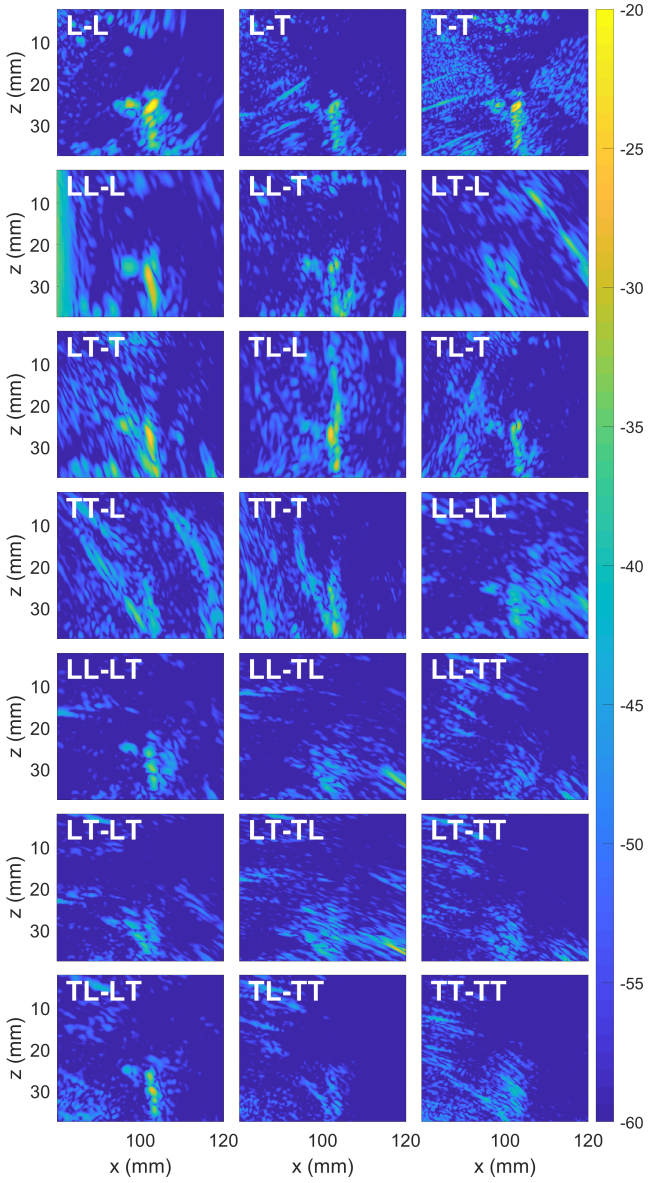

Fig. 11. Multi-view TFM images (in dB relative to the peak back wall response in the L-L view) of defect a (at $y = 57$ mm) within Sample B.

on the likelihood of not being noise, i.e. novelty detection. Methodologies for combining p-values have been widely studied [21], [22]. However, the limitation of equating p-values based on a noise response rather than a signal response introduces complexities when sizing the defects themselves as established sizing methodologies can no longer be used and thus is not examined in this work.

VI. Conclusion

Maximising the visualisation of large defects through data fusion of multi-view TFM images has been outlined. Three data fusion methodologies were examined, namely the naive sum, the micro-structural noise-weighted sum and the matched filter. In the first method, no prior knowledge is utilised when combining the data. In the second, only micro-structural noise information is included whilst the matched filter uses a priori knowledge on the expected signal response together with the noise information. Increasing the foreknowledge was shown to provide meaningful benefit, with the matched filter providing the best results, with significant improvements over a single view for three of the four defects examined. Whilst a priori knowledge of a defect's signal response can be tailored to a specific defect type and orientation, this work utilised an analytical model for a small radius (0.001mm) SDH, representative of a generic omni-directional point scatterer whilst maintaining relative scattering between the LL, LT, TL and TT inbound-outbound modalities from the scatterer. This enabled each TFM view to be combined on a pixel-by-pixel basis, weighted by its expected SNR in each view. To avoid non-defect signals contaminating the fused image, a masking process, based on a single defect-free training dataset, was also employed for the two fusion methodologies utilising noise information. This mask is a useful by-product of the micro-structural noise quantification procedure.

Obtaining the necessary *a priori* knowledge for the matched filter required only a single dataset for the noise parameters and a single simulation for the expected signal, per experimental setup. Even for the second sample, which demonstrated

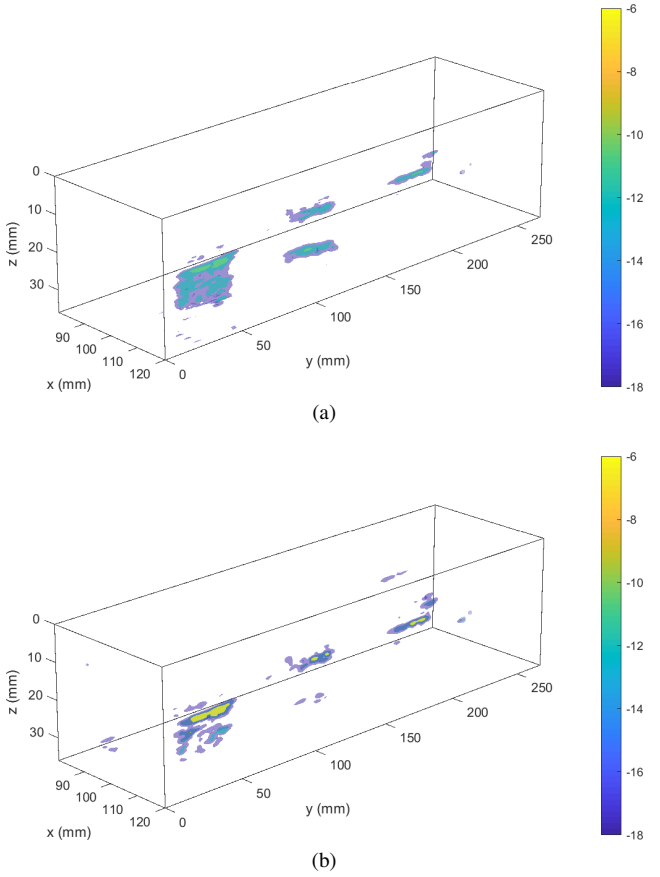

Fig. 12. 3D images of Sample B interior: (a) fused result using matched filter; (b) L-L view. Isosurfaces are plotted at -6, -12 and -18dB relative to the peak defect response in each image.

irregular surface geometry and a 3% thickness variation, a single dataset was sufficient, with the biggest (yet still minor) impact evident on the mask effectiveness for defect *c* of Sample B.

The proposed method has been described as a characterisation rather than a detection tool. However, since the fused image can be generated rapidly from FMC data, it may also be used for defect detection. An obvious practical example is weld inspection in a configuration similar to that used for Sample B. Here the array is configured so that the ROI encompasses the complete weld cross-section and the fused multi-view images from a scan provide complete volumetric coverage of the weld. This means that the same data can be used for both the detection of defects and their subsequent characterisation. This is particularly relevant in cases such as weld inspection, where an oblique incidence scan is often performed along the length of the weld anyway.

Image registration was controlled in this work through determination of the experimental setup parameters, e.g. longitudinal specimen velocity, using the training dataset rather than relying on book values. For the second sample, it was also necessary to accurately measure the front and back surfaces prior to forming the multi-view TFM images. Without accounting for this surface geometry there was a significant mislocation of defect responses and it was observed that this degraded the matched filter considerably. This impact varied

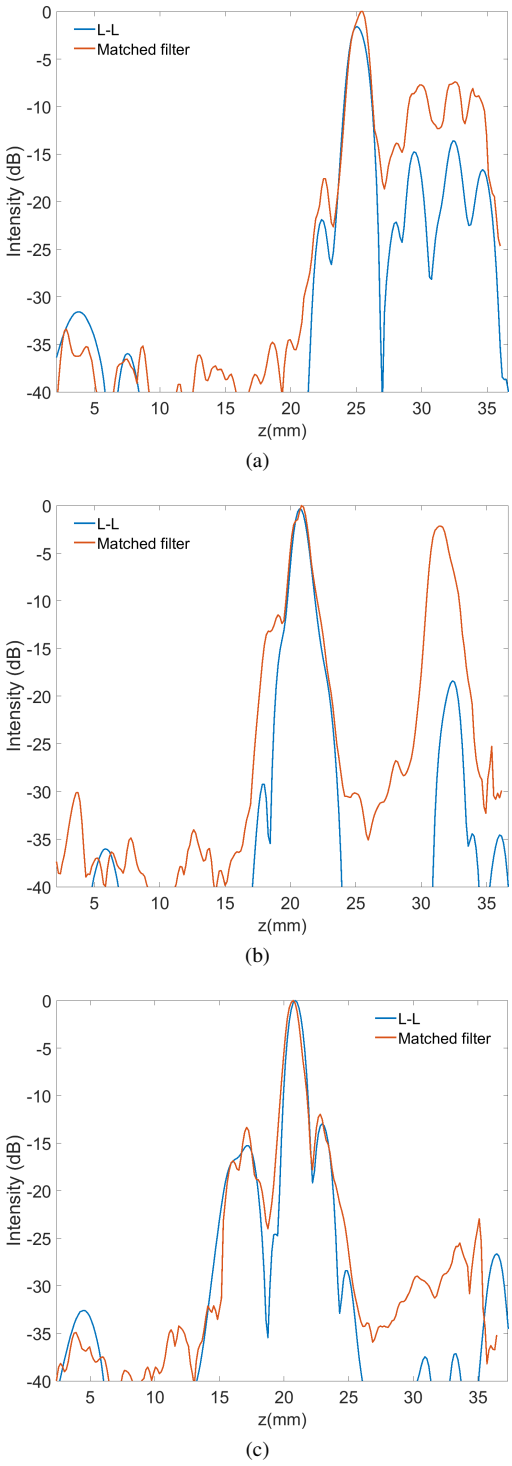

Fig. 13. Intensity profiles along defect vertical centreline in Sample B: (a) defect *a*; (b) defect *b*; (c) defect *c*. Each profile is normalised independently by its respective peak intensity.

throughout the imaged region of interest and primarily affected the two weld crack defects in this case due to the weaker specular components.

The matched filter approach to data fusion outlined in this paper is shown to be particularly useful when dealing with large defects, as the whole of the defect extent is typically not fully visible from one particular imaging modality, as shown for the SCC in Sample A, where one fork of the V shaped crack produced a weaker response (figure 7). Overall, data fusion provides a systematic and robust way of combining information from multiple views into a single image that provides a better representation of the true extent of a defect.

Data Access Statement

All data used in this paper are openly available for download from the University of Bristol Research Data Repository at [DOI WILL BE ADDED AFTER REVIEW]

References

[1] B. W. Drinkwater and P. D. Wilcox, "Ultrasonic arrays for non-destructive evaluation: A review," *NDT & E International*, vol. 39, no. 7, pp. 525–541, 10 2006.

[2] C. Holmes, B. W. Drinkwater, and P. D. Wilcox, "Post-processing of the full matrix of ultrasonic transmit-receive array data for non-destructive evaluation," *NDT & E International*, vol. 38, no. 8, pp. 701–711, 2005.

[3] A. J. Hunter, B. W. Drinkwater, and P. D. Wilcox, "The wavenumber algorithm for full-matrix imaging using an ultrasonic array," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 55, no. 11, pp. 2450–2462, 11 2008.

[4] J. Zhang, B. W. Drinkwater, and P. D. Wilcox, "Efficient immersion imaging of components with non-planar surfaces," *IEEE Trans. Ultrason. Ferr. Freq. Contr.*, vol. 61, no. 8, pp. 1284–1295, 2014.

[5] J. Zhang, T. Barber, A. Nixon, and P. Wilcox, "Investigation into distinguishing between small volumetric and crack-like defects using multi-view total focusing method images," *AIP conference proceedings*, vol. 1806, p. 040003, 2017.

[6] J. Zhang, B. W. Drinkwater, P. D. Wilcox, and A. J. Hunter, "Defect detection using ultrasonic arrays: The multi-mode total focusing method," *NDT & E International*, vol. 43, no. 2, pp. 123 – 133, 2010.

[7] L. L. Jeune, S. Robert, E. L. Villaverde, and C. Prada, "Plane wave imaging for ultrasonic non-destructive testing: Generalization to multi-modal imaging," *Ultrasonics*, vol. 64, pp. 128–138, 2016.

[8] L. Merabet, S. Robert, and C. Prada, "The multi-mode plane wave imaging in the fourier domain: Theory and applications to fast ultrasound imaging of cracks," *NDT & E International*, vol. 110, p. 102171, 2020.

[9] K. Sy, P. Bredif, E. Iakovleva, O. Roy, and D. Lesselier, "Development of the specular echoes estimator to predict relevant modes for total focusing method imaging," *NDT & E International*, vol. 99, pp. 134–140, 2018.

[10] Z. Liu, D. S. Forsyth, J. P. Komorowski, K. Hanasaki, and T. Kirubarajan, "Survey: State of the art in NDE data fusion techniques," *IEEE Transactions On Instrumentation And Measurement*, vol. 56, no. 6, pp. 2435–2451, 2007.

[11] N. Brierley, T. Tippetts, and P. Cawley, "Data fusion for automated non-destructive inspection," *Proceedings of the Royal Society A*, vol. 470, no. 2167, 2014.

[12] R. F. Wagner, S. W. Smith, J. M. Sandrik, and H. Lopez, "Statistics of speckle in ultrasound b-scans," *IEEE Transactions On Sonics And Ultrasonics*, vol. 30, pp. 156–163, 1983.

[13] R. L. T. Bevan, J. Zhang, N. Budyn, A. J. Croxford, and P. D. Wilcox, "Experimental quantification of noise in linear ultrasonic imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 66, no. 1, pp. 79–90, 2019.

[14] J. Lubbers and R. Graaff, "A simple and accurate formula for the sound velocity in water," *Ultrasound in Medicine and Biology*, vol. 24, no. 7, pp. 1065–1068, 1998.

[15] Zhen Ye, Jasjit Suri, Yajie Sun, and R. Janer, "Four image interpolation techniques for ultrasound breast phantom data acquired using fischer's full field digital mammography and ultrasound system (ffdmus): a comparative approach," in *IEEE International Conference on Image Processing 2005*, vol. 2, Sep. 2005, pp. II–1238.

[16] N. Budyn, R. L. T. Bevan, J. Zhang, A. J. Croxford, and P. D. Wilcox, "A model for multi-view ultrasonic array inspection of small two-dimensional defects," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 66, no. 6, pp. 1129–1139, 2019.

[17] E. Iakovleva, S. Chatillon, P. Bredif, and S. Mahaut, "Multi-mode TFM imaging with artefacts filtering using CIVA UT forwards models," *40th Annual Review of Progress in Quantitative Nondestructive Evaluation*, vol. 1581, pp. 72–79, 2014.

[18] A. L. Lopez-Sanchez, H.-J. Kim, L. W. Schmerr, and A. Sedov, "Measurement models and scattering models for predicting the ultrasonic pulse-echo response from side-drilled holes," *Journal of Nondestructive Evaluation*, vol. 24, no. 3, pp. 83–96, Sep 2005.

[19] R. Murphy, "Ultrasonic defect-sizing using decibel drop methods volume 1: text," Atomic Energy Control Board, Tech. Rep. INFO-0250-1, 03 1987.

[20] J. G. McKee, P. D. Wilcox, and R. E. Malkin, "Effect of surface compensation for imaging through doubly-curved surfaces using a 2d phased array," in *AIP Conference Proceedings*, vol. 2102, no. 1. AIP Publishing, 2019, p. 100008.

[21] A. Birnbaum, "Combining independent tests of significance," *Journal of the American Statistical Association*, vol. 49, pp. 559–574, 1954.

[22] A. M. Winkler, M. A. Webster, J. C. Brooks, I. Tracey, S. M. Smith, and T. E. Nichols, "Non-parametric combination and related permutation tests for neuroimaging," *Human Brain Mapping*, vol. 37, pp. 1486–1511, 2016.